Higher order asymptotic optimality in testing problems with nuisance parameters

Vladimir E. Bening[1], Dimitrii M. Chibisov[2]

January 15, 1998

Math. Sub Class.: 62F05

Key words: Hypothesis testing, nuisance parameters, asymptotic methods, deficiency

[1]Research supported by the Grant INTAS-RFBR 95-0099 and by the Grant 047-003-012 from the Netherlands Organization of Scientific Research.

[2]Research supported by the Grant 96-15-96033 from the Russian Foundation for Fundamental Research, INTAS Grants 93-1585, 94-4194, and INTAS-RFBR 95-0099, by the Grant 047-003-012 from the Netherlands Organization of Scientific Research, and by the Weierstrass Institute of Applied Analysis and Stochastics.

Abstract

We consider testing hypotheses about the location parameter of a symmetric distribution when a finite-dimensional nuisance parameter is present. For local alternatives, we study the power loss of asymptotically efficient tests in this problem, which is the difference between the power of the most powerful test for a given value of the nuisance parameter (as if it were known) and the power of the test at hand. The power loss is typically of order n^{-1} and is closely related to the deficiency of the test. In particular, we obtain the lower bound for the power loss in a locally asymptotically minimax sense similar to that used in the estimation theory and indicate a test on which this bound is attained. This bound corresponds to the envelope power function obtained by Pfanzagl and Wefelmeyer (1978) for test statistics of a specific structure.

1 Introduction

In this paper we study asymptotically efficient tests for hypotheses about a univariate parameter when a finite-dimensional nuisance parameter is present. This problem was investigated by Pfanzagl and Wefelmeyer (1978) (see also the review paper Pfanzagl (1980)) who described asymptotically complete classes of tests in this setting. We obtain related results in a technically simpler way. To simplify the presentation, we treat the case where the underlying distribution is symmetric about the location parameter of interest. The main tool is a formula for the difference between the powers of the most powerful (MP) test for a simple hypothesis against a simple (local) alternative and an asymptotically efficient test in the same testing problem. Using this formula we do not derive asymptotic expansions for the powers of tests, dealing directly with the power loss of tests for the composite hypothesis as compared to the MP test for the case the nuisance parameter were known. We obtain lower bounds for this power loss and indicate tests on which they are attained.

Specifically, we consider testing the hypothesis

$$H_0 : \theta = \theta_0, \zeta \in Z \quad \text{against} \quad H_1 : \theta > \theta_0, \zeta \in Z$$

based on i.i.d. real-valued observations $X_1, \dots, X_n$ with symmetric Lebesgue density

$$p_{\theta,\zeta}(x) = p_\zeta(x - \theta), \qquad p_\zeta(x) = p_\zeta(-x),$$

where $\zeta = (\zeta_1, \dots, \zeta_k) \in Z$ with an open $Z \subset R^k$. Henceforth without loss of generality we take $\theta_0 = 0$. We generically denote by $\beta_n(t, \eta)$ the power of a test for H_0 against the local alternative $(n^{-1/2}t, \eta)$, $t > 0$. In particular, $\beta_n(0, \eta)$ is the test size, and we restrict ourselves to asymptotically (as.) similar tests satisfying, for a fixed level $\alpha > 0$, the condition

$$\sup_{\zeta \in K} |\beta_n(0, \zeta) - \alpha| = o(n^{-1})$$

for any compact subset $K \subset Z$.

An immediate way of obtaining an upper bound for the power of an arbitrary as. similar test is as follows. Consider testing a simple hypothesis $(0, \zeta)$ against a simple alternative

$(n^{-1/2}t, \eta)$, $\zeta, \eta \in Z$. Let $\beta_n(t, \eta; \zeta)$ be the power of the MP size α test in this testing problem. Then the power $\beta_n(t, \eta)$ of any as. similar test is no greater than $\beta_n(t, \eta; \zeta) + o(n^{-1})$ for any $\zeta \in Z$. Hence

$$\beta_n(t, \eta) \leq \bar{\beta}_n(t, \eta) + o(n^{-1}), \tag{1.1}$$

where

$$\bar{\beta}_n(t, \eta) = \inf_{\zeta} \beta_n(t, \eta; \zeta).$$

The minimizer of $\beta_n(t, \eta; \zeta)$ is the least favorable hypothesis for the given alternative. This bound was derived by Pfanzagl and Wefelmeyer (1978) and Pfanzagl (1980) ($E_2^{*(n)}$ in their notation, see (10.2.4)). They do not restrict themselves to symmetric distributions. In this general case the least favorable hypothesis is randomized (see Pfanzagl (1980), p. 50). In our special case this randomization is not needed.

In contrast to Pfanzagl and Wefelmeyer (1978) we do not derive an asymptotic expansion for $\bar{\beta}_n(t, \eta)$. Put $\beta_n^*(t, \eta) = \beta_n(t, \eta; \eta)$, so that $\beta_n^*(t, \eta)$ is the power of the MP test for $(0, \eta)$ against $(n^{-1/2}t, \eta)$, which could be achieved if the nuisance parameter η were known. Without deriving separately asymptotic expansions for $\beta_n^*(t, \eta)$ and $\bar{\beta}_n(t, \eta)$ we directly obtain an asymptotic formula of the form

$$\beta_n^*(t, \eta) - \bar{\beta}_n(t, \eta) = n^{-1}B(t, \eta) + o(n^{-1}) \tag{1.2}$$

for their difference (see (3.24) or (3.26)).

Using this formula the asymptotic expansion for $\bar{\beta}_n$ can be immediately derived from the well-known asymptotic expansion for $\beta_n^*(t, \eta)$ (see, e.g., Pfanzagl (1980), (9.4.1)). However the n^{-1} term of the difference $\beta_n^*(t, \eta) - \beta_n(t, \eta)$ determines the *deficiency* of the corresponding test (see, e.g., Pfanzagl (1980), p. 73), so that this difference is of interest in its own right. We refer to such a difference as the *power loss* of the test and deal with power losses of tests rather than deriving corresponding deficiencies.

It is seen from (1.1) and (1.2) that the RHS of (1.2) provides a lower bound for the power loss of an arbitrary as. similar test for H_0. In Section 3.2 we construct as. similar tests on which this bound is attained. However these tests depend on the chosen parameter point η and the lower bound is attained in a small neighborhood of η. This resembles the superefficiency effect in estimation, where a lower risk than the regular (Cramér–Rao) bound can be attained in a vicinity of a given parameter point at the expense of increase of the risk elsewhere. This suggests the local minimax approach characterizing a test by the maximal loss over a small neighborhood in the parameter space.

Denote by $\mathcal{S}$ the class of as. similar size α tests and by $\beta_n^\phi(t, \zeta)$ the power of a test $\phi \in \mathcal{S}$ at the alternative $(n^{-1/2}t, \zeta)$. It will be expedient here to normalize the deviation of θ by $\sqrt{nJ_\zeta}$ rather than $\sqrt{n}$, where J_ζ is the Fisher information w.r.t. θ for fixed $\zeta \in Z$. With this normalization the powers under consideration converge to a limit depending only on t, but not on the nuisance parameter.

Thus for given $t > 0$ and $\eta \in Z$ the power loss of a test $\phi \in \mathcal{S}$ at an alternative $(t/\sqrt{nJ_\eta}, \eta)$ is characterized by

$$r_n^\phi(t, K) = \sup_{\zeta \in K} (\beta_n^*(t/\sqrt{J_\zeta}, \zeta) - \beta_n(t/\sqrt{J_\zeta}, \zeta)),$$

where $K \subset Z$ is a neighborhood of η, and $\beta_n^*(t/\sqrt{J_\zeta}, \zeta)$, as before, is the power of the MP test for $(0, \zeta)$ against $((nJ_\zeta)^{-1/2}t, \zeta)$. We establish an asymptotic lower bound for $nr_n^\phi(t, K)$ as $n \to \infty$ and K shrinks to η, i.e., we show that

$$\lim_{K \downarrow \eta} \liminf_{n \to \infty} \inf_{\phi \in S} nr_n^\phi(t, K) \geq B^*(t, \eta), \tag{1.3}$$

where $B^*(t, \eta)$ is given by (4.31). As is to be expected, this bound is no less and, in general, greater than the bound (1.2) corresponding to the least favorable hypothesis. In **4.2** we construct tests attaining this bound for any value of the nuisance parameter (uniformly on compact sets). The corresponding upper bound for the power was obtained by Pfanzagl and Wefelmeyer (1978) (see p. 57, Theorem 2) in the context of deriving an as. complete class of tests. They call it *the envelope power function*. The properties stated above justify this term.

The treatment in Pfanzagl and Wefelmeyer (1978) is restricted to test statistics of certain structure (admitting a stochastic expansion). We do not impose any restrictions on the tests under consideration. The particular form of the family of distributions (symmetric distributions with a location parameter of interest) was adopted to work out the approach and techniques in a simplified setting. More general families can be treated along the same lines.

This study was motivated by the problem of calculating deficiencies of asymptotically efficient adaptive tests in a semiparametric setup. In the simplest case this problem is as follows. We want to test hypotheses about the real-valued location parameter θ given i.i.d. observations $X_1, \ldots, X_n$ with Lebesgue density $p(x - \theta)$, where $p(x)$ is an unknown density symmetric about zero, $p(x) = p(-x)$. Suppose we are testing a simple hypothesis against one-sided alternatives, viz,

$$H_0 : \theta = 0 \quad \text{against} \quad H_1 : \theta > 0.$$

When p is known (and satisfies certain regularity conditions), the MP test against a local alternative of the form $\theta = tn^{-1/2}$, $t > 0$, has a nontrivial power $\beta_n^*(t, p)$ bounded away from α and one. For unknown (symmetric) p one can construct adaptive tests having asymptotically the same power, i.e., the power $\beta_n(t, p)$ such that

$$\beta_n^*(t, p) - \beta_n(t, p) \to 0 \qquad \text{as} \quad n \to \infty.$$

It is natural to ask about the rate of this convergence. More precisely, like it is done in estimation problems, to look for a lower bound for $\beta_n^* - \beta_n$ and, if possible, to construct tests attaining this bound, which would then be (higher-order) asymptotically efficient.

In this setting the density p can be viewed as an infinite-dimensional nuisance parameter. The present paper is an attempt to find an approach in the finite-dimensional setup, which could be extended to the infinite-dimensional case.

This paper is written in an informal style. We do not state regularity conditions and do not give formal proofs. Rather, we try to demonstrate in the most transparent way how the results can be derived. The formal proofs will be given in a subsequent paper.

We begin with the case of no nuisance parameter (Section 2). This case is presented to introduce in the simplest possible setting some notions and results which are then used in the nuisance parameter setup. In Section 3 we derive the bound (1.2) related to the least favorable hypothesis and indicate a test attaining this bound. As we pointed out, this test depends on the chosen value of the nuisance parameter. The locally asymptotically minimax bound (1.3) and a test attaining this bound are constructed in Section 4. Section 5, Appendix, contains informal proofs of some auxiliary results.

2 No nuisance parameter case

2.1 LLR and first-order efficiency

We have i.i.d. observations $X_1, \ldots, X_n$ with density $p_\theta(x)$. In this case we need not assume θ to be a location parameter (which is assumed in the nuisance parameter setup for some simplification), so that X's can take values in an arbitrary measurable space $(\mathcal{X}, \mathcal{A})$ and p_θ is their common density function w.r.t. some σ-finite measure on $\mathcal{X}$. We test the hypothesis

$$H_0 : \theta = 0 \quad \text{against} \quad H_1 : \theta > 0. \tag{2.1}$$

Throughout the paper we use the abbreviation

$$\tau = n^{-1/2}. \tag{2.2}$$

For any $t > 0$ we will also consider the simple alternative

$$H_{n,t} : \theta = \tau t. \tag{2.3}$$

We denote by $P_{n,0}$ and $P_{n,t}$ the joint distributions of $\mathbf{X} = (X_1, \ldots, X_n)$ under H_0 and $H_{n,t}$ respectively. Obviously, they have densities

$$p_{n,0}(\mathbf{x}) = \prod_1^n p_0(x_i) \quad \text{and} \quad p_{n,t}(\mathbf{x}) = \prod_1^n p_{\tau t}(x_i) \tag{2.4}$$

w.r.t. the corresponding product measure, $\mathbf{x} = (x_1, \ldots, x_n)$. The respective expectations will be denoted by $E_{n,0}$ and $E_{n,t}$ (with subscript n dropped when applied to a function of a single X).

Assume that all measures $P_{n,t}$ are mutually absolutely continuous. Consider the loglikelihood ratio (LLR)

$$\Lambda_n(t) = \log \frac{dP_{n,t}}{dP_{n,0}} = \log \frac{p_{n,t}}{p_{n,0}}. \tag{2.5}$$

We denote $l(x) = \log p(x)$ with corresponding indices. Then by (2.4)

$$\Lambda_n(t) = \sum [l_{\tau t}(X_i) - l_0(X_i)]. \tag{2.6}$$

By the Taylor series expansion,

$$l_{\tau t}(X_i) - l_0(X_i) = \tau t l^0(X_i) + \frac{1}{2}(\tau t)^2 l^{00}(X_i) + \ldots \tag{2.7}$$

Here and in what follows we denote by the superscript 0 the differentiation w.r.t. θ. (When the nuisance parameter is present, the differentiation w.r.t. its ith component will be denoted by the superscript i.) We omit the subscript when the derivative is taken at $\theta = 0$. Denote

$$L_n^0 = \tau \sum l^0(X_i), \quad L_n^{00} = \tau \sum [l^{00}(X_i) - E_0 l^{00}], \dots \tag{2.8}$$

The sums are centered by the corresponding E_0-expectations; the first sum contains no centering because $E_0 l^0 = 0$. Further, denote by J the Fisher information

$$J = E_0(l^0)^2. \tag{2.9}$$

(We reserve the usual notation I for the Fisher information w.r.t. the nuisance parameter.) It is well known that $E_0 l^{00} = -J$. With this notation, putting (2.7) into (2.6) yields

$$\Lambda_n(t) = tL_n^0 - \frac{1}{2}t^2 J + \frac{1}{2}\tau t^2 L_n^{00} + \dots \tag{2.10}$$

The first two terms in the RHS of (2.10) express the local asymptotic normality (LAN) of the family of distributions. The omitted terms include the nonrandom term $\frac{1}{6}\tau t^3 E_0 l^{000}$ and the terms of higher order than τ.

The most powerful (MP) size α test for H_0 against $H_{n,t}$ rejects H_0 when $\Lambda_n(t) > c_{n,t}$ with $c_{n,t}$ defined by

$$P_{n,0}(\Lambda_n(t) > c_{n,t}) = \alpha. \tag{2.11}$$

(We tacitly assume continuity of the corresponding distribution.) By the CLT

$$\mathcal{L}(\Lambda_n(t) \mid P_{n,0}) \to N(-\frac{1}{2}t^2 J, t^2 J). \tag{2.12}$$

Hence

$$c_{n,t} \to c_t = t\sqrt{J}u_{1-\alpha} - \frac{1}{2}t^2 J, \tag{2.13}$$

$u_{1-\alpha}$ denoting the upper α-point of the standard normal distribution. The power of this MP test is

$$\beta_n^*(t) = P_{n,t}(\Lambda_n(t) > c_{n,t}). \tag{2.14}$$

It is known from the LAN theory that

$$\mathcal{L}(\Lambda_n(t) \mid P_{n,t}) \to N(\frac{1}{2}t^2 J, t^2 J). \tag{2.15}$$

Thus (2.13)–(2.15) yield

$$\beta_n^*(t) \to \beta(t) = \Phi(t\sqrt{J} - u_{1-\alpha}), \tag{2.16}$$

where Φ stands for the standard normal d.f. and $\Phi(u_{1-\alpha}) = 1 - \alpha$.

Note that $\beta_n^*(t)$, known as the *envelope power function*, is not the power function of a single test. For each $t > 0$ it is the power of the MP test against $H_{n,t}$ based on $\Lambda_n(t)$. Thus it provides an upper bound for the power of any test for H_0 against $H_1 : t > 0$.

It is well known that there are many (first order) asymptotically efficient tests, i.e., tests whose power function $\beta_n(t)$ converges to the same limit as $\beta_n^*(t)$. So are, for example, tests based on L_n^0, on $\Lambda_n(t_0)$ with an arbitrary $t_0 > 0$, on the MLE $\hat{\theta}_n$, on a certain linear combination of order statistics; for θ location parameter there are asymptotically efficient rank tests. They can be compared with each other by higher order terms of their power.

Before proceeding to the higher-order theory, we will derive some simple formulas to be used in the sequel.

Denote by $f_{0,t}(x)$ and $f_{1,t}(x)$ the limiting densities of $\Lambda_n(t)$ under $P_{n,0}$ and $P_{n,t}$ respectively, which correspond to the normal distributions in (2.12) and (2.15). Note that they are related to each other by $e^x f_{0,t}(x) = f_{1,t}(x)$, which follows from the properties of the LLR or can be verified directly. We will need expressions for $f_{0,t}(c_t)$ and $f_{1,t}(c_t)$. Putting (2.13) into the explicit expressions for normal densities (2.12), (2.15) yields

$$f_{0,t}(c_t) = \frac{1}{t\sqrt{J}}\varphi(u_{1-\alpha}), \quad f_{1,t}(c_t) = \frac{1}{t\sqrt{J}}\varphi(u_{1-\alpha} - t\sqrt{J}). \tag{2.17}$$

Next, suppose instead of $c_{n,t}$ we use another critical value $c'_{n,t}$, say, which also converges to c_t (see (2.12), (2.13)). Then the test $\Lambda_n(t) > c'_{n,t}$ has size α'_n and power $\beta^{*'}_n(t)$ converging to α and $\beta(t)$ respectively. Let us now have two such sequences $c'_{n,t}$ and $c''_{n,t}$ converging to c_t with $\delta_n = c'_{n,t} - c''_{n,t} \to 0$, and we are interested in the differences of the corresponding sizes and powers up to $o(\delta_n)$. Assuming certain regularity, so that the d.f.'s of $\Lambda_n(t)$ under $P_{n,0}$ and $P_{n,t}$ have Edgeworth expansions, it is easy to see that these differences are entirely determined by the leading terms of these expansions, because the next terms contribute at most $O(\tau\delta_n) = o(\delta_n)$. The leading terms are the normal distributions we have just discussed. Thus it is readily seen that

$$\alpha''_n - \alpha'_n = \delta_n f_{0,t}(c_t) + o(\delta_n) = \frac{\delta_n}{t\sqrt{J}}\varphi(u_{1-\alpha}) + o(\delta_n), \tag{2.18}$$

$$\beta^{*''}_n(t) - \beta^{*'}_n(t) = \delta_n f_{1,t}(c_t) + o(\delta_n) = \frac{\delta_n}{t\sqrt{J}}\varphi(u_{1-\alpha} - t\sqrt{J}) + o(\delta_n). \tag{2.19}$$

2.2 Second order efficiency

Typically, an asymptotically efficient test statistic (suitably normalized) has the score function L^0_n as its leading term, so that it has the form

$$T_n = L^0_n + \tau H_n + \ldots, \tag{2.20}$$

with H_n bounded in probability. For example (see (2.10)) $\Lambda_n(t_0)$ is equivalent to $T_n = L^0_n + \frac{1}{2}\tau t_0 L^{00}_n$. For rank statistics and linear combinations of order statistics H_n can be written as a quadratic functional of the empirical process (centered and normalized empirical d.f.).

In 70-ies expansions in τ to terms of order τ^2 were obtained for the power functions $\beta_n(t)$ of various asymptotically efficient tests. The purpose was to study the deficiencies of the corresponding tests, which we will briefly discuss later on. Writing down such expansions in an explicit form required very involved calculations. For "parametric" test statistics first a "stochastic expansion" of the form (2.20), but containing also the τ^2 term was derived. It was used to obtain the Edgeworth expansions (briefly, E-expansions) for the distributions of T_n under $P_{n,0}$ and $P_{n,t}$. (For rank statistics a different technique based on a certain conditioning was used by Albers, Bickel, and van Zwet (1976).) The

E-expansion under $P_{n,0}$ was used to obtain an expansion in τ for the critical value a_n defined by $P_{n,0}(T_n > a_n) = \alpha$. Then the E-expansion for

$$\beta_n(t) = P_{n,t}(T_n > a_n)$$

was derived by the substitution of the expansion for a_n into the E-expansion under $P_{n,t}$.

Though the E-expansions for the distributions of various asymptotically efficient test statistics and of $\Lambda_n(t)$ differ by terms of order τ, it was observed that their powers $\beta_n(t)$ differ from each other and from $\beta_n^*(t)$ by $o(\tau)$ (and typically by $O(\tau^2)$), so that "first-order efficiency implies second-order efficiency", the latter meaning that the power agrees with $\beta_n^*(t)$ up to terms of order τ. The approach of comparing the expansions for β_n^* and β_n described above gave no insight into the nature of this phenomenon. A simple and intuitively clear proof of this general property was given by Bickel, Chibisov, and van Zwet (1981). We outline here that proof adapted to the present setup.

The idea was, first, to treat directly the difference $\beta_n^*(t) - \beta_n(t)$ and, secondly, to adjust the test statistic to the LLR (rather than to adjust test statistics and the LLR to L_n^0), so that the difference

$$\Delta_{n,t} := \Lambda_n(t) - S_{n,t} \tag{2.21}$$

is small. For example, (2.20) as a test statistic is equivalent to

$$S_{n,t} = tT_n - \frac{1}{2}t^2 J$$

and then (see (2.10))

$$\Delta_{n,t} = \tau(\frac{1}{2}\tau t^2 L_n^{00} - tH_n) + \dots$$

(We state this expression to show that $\Delta_{n,t}$ is of order τ and do not need its particular form.) Throughout the rest of this section we mostly suppress the subscript and argument t. Let c_n and b_n be the corresponding critical values defined by

$$P_{n,0}(\Lambda_n > c_n) = P_{n,0}(S_n > b_n) = \alpha. \tag{2.22}$$

Then the corresponding powers are

$$\beta_n^* = P_{n,t}(\Lambda_n > c_n), \quad \beta_n = P_{n,t}(S_n > b_n).$$

Their difference is

$$\beta_n^* - \beta_n = \int_{\{\Lambda_n > c_n\}} dP_{n,t} - \int_{\{S_n > b_n\}} dP_{n,t} = \int_{A_+} dP_{n,t} - \int_{A_-} dP_{n,t}, \tag{2.23}$$

where

$$A_+ = \{\Lambda_n > c_n, S_n \le b_n\}, \quad A_- = \{\Lambda_n \le c_n, S_n > b_n\}. \tag{2.24}$$

Since $dP_{n,t} = e^{\Lambda_n} dP_{n,0}$ and

$$\int_{\{\Lambda_n > c_n\}} dP_{n,0} - \int_{\{S_n > b_n\}} dP_{n,0} = 0$$

by (2.22), we can rewrite (2.23) as

$$\beta_n^* - \beta_n = \Big(\int_{A_+} - \int_{A_-}\Big)(e^{\Lambda_n} - e^{c_n})dP_{n,0}. \tag{2.25}$$

Using (2.21) rewrite (2.24) as

$$A_+ = \{c_n < \Lambda_n \le b_n - \Delta_n\}, \quad A_- = \{b_n - \Delta_n < \Lambda_n \le c_n\}. \tag{2.26}$$

Since Δ_n is of order τ, so is the difference of d.f.'s of Λ_n and S_n, hence so is $c_n - b_n$. Thus Λ_n in (2.25) varies in the layer (2.26) having width of order τ. Moreover, the integrand in (2.25) vanishes on one side of this layer, namely, on the surface $\Lambda_n = c_n$, so that it remains $O(\tau)$ in the domain of integration. Its integration over the thin layer results in

$$\beta_n^* - \beta_n = o(\tau). \tag{2.27}$$

An argument of this type was used in Bickel, Chibisov, and van Zwet (1981) to obtain (2.27) under very general conditions, in particular, on the magintude of Δ. When Δ is of order τ, it is seen from the above argument that the difference in (2.27) is likely to be $O(\tau^2)$.

2.3 Power loss

The difference $\beta_n^*(t) - \beta_n(t)$ is closely related to the *deficiency* of the corresponding test, which is the number of additional observations needed for this test to achieve the same power as the MP test. This notion was introduced by Hodges and Lehmann (1970). Deficiencies of various tests were extensively studied in 70-ies by Albers, Bickel, and van Zwet (1976) (for rank tests), by Chibisov (1983), Pfanzagl (1980) (for "parametric" tests) and others. When the limit

$$B(t) := \lim_n n(\beta_n^*(t) - \beta_n(t)) \tag{2.26}$$

exists, the asymptotic deficiency is finite and can be directly expressed through this limit. We will not state this relationship here. Rather, we will directly deal with the quantity (2.26), which we will refer to as the *power loss*. This quantity was actually the object of the studies on deficiency. As we pointed out, its derivation was very involved.

An elaboration of the argument given in the previous subsection leads to the following formula for the power loss. Suppose that Δ_n as in (2.21) is of order τ in a somewhat stronger sense then it was meant before. Namely, assume that $(\sqrt{n}\Delta_n, \Lambda_n)$ converges in distribution under $P_{n,0}$ to a certain bivariate r.v. Denoting $\Pi_n = \sqrt{n}\Delta_n$, we write it as

$$(\Pi_n, \Lambda_n) \xrightarrow{P_{n,0}} (\Pi, \Lambda). \tag{2.27}$$

In all regular cases Λ is a normal r.v. (see (2.12)). Denote its d.f. and density by $F_0(x)$ and $f_0(x)$. Let c be the limiting critical value defined by $F_0(c) = 1 - \alpha$. Then

$$\lim n(\beta_n^* - \beta_n) = \frac{1}{2} e^c f_0(c) \operatorname{Var}[\Pi \mid \Lambda = c]. \tag{2.28}$$

(Note that $e^c f_0(c) = f_1(c)$, where f_1 is the limiting density of Λ_n under $P_{n,t}$.) In the above argument we assumed that the tests have exactly size α (see (2.22)), but the formula (2.28) remains valid when the sizes converge to α and equal each other up to $o(\tau^2)$, i.e.,

$$P_{n,0}(\Lambda_n > c_n) - P_{n,0}(S_n > b_n) = o(\tau^2). \tag{2.29}$$

The formula (2.28) demonstrates, in particular, that the power loss (hence the deficiency) is determined by the terms of order τ of the (asymptotically efficient) test statistic.

We give an informal proof of (2.28) in **5.3**. This "proof" was first presented in Chibisov (1982). Its justification, however, depends on the structure of T_n. The formula (2.28) was proved by Chibisov (1985) for statistics admitting a stochastic expansion in terms of i.i.d. sums (which is typical for "parametric" problems and applicable in the setting of this paper). Bening (1995, 1997) proved formula (2.28) for rank statistics, linear combinations of order statistics and U-statistics.

3 Least favorable hypothesis

3.1 Local lower bound for the power loss

Now we consider i.i.d. real-valued observations $X_1, \ldots, X_n$ with density

$$p_{\theta,\zeta}(x) = p_\zeta(x - \theta), \tag{3.1}$$

where $\zeta = (\zeta_1, \ldots, \zeta_k) \in Z \subset R^k$ and $p_\zeta(x) = p_\zeta(-x)$ for all ζ. We test the hypothesis

$$H_0 : \theta = 0, \zeta \in Z \quad \text{against} \quad H_1 : \theta > 0, \zeta \in Z. \tag{3.2}$$

The main distinction from the no nuisance parameter case is that the test size depends now on unknown ζ. Thus we will look for as. similar tests with size approximately equal to α in some asymptotic sense.

For some $t > 0$ and $\eta \in Z$ (not to write ζ with additional indices like ζ_0) consider the sequence of simple local alternatives

$$H_{n,t,\eta} : (\theta, \zeta) = (\tau t, \eta). \tag{3.3}$$

If η were known, we would have a location family $p_\eta(x - \theta)$, which was considered in the previous section. Denote by $\beta_n^*(t, \eta)$ the power of the MP test for $(0, \eta)$ against $(\tau t, \eta)$. Since any as. efficient as. similar test for the problem (3.2) is an as. efficient test for this testing problem, we can evaluate the difference between its power, $\beta_n(t, \eta)$, say, and $\beta_n^*(t, \eta)$ by the formula (2.28). As we pointed out, such difference is treated much easier than the power itself.

We will consider as. similar tests which have size $\alpha + o(\tau^2)$ uniformly over compact subsets of Z. This requirement can be written as

$$\sup_{\zeta \in K} |\beta_n(0, \zeta) - \alpha| = o(\tau^2) \tag{3.4}$$

for any compact set $K \subset Z$. A lower bound for the power loss of any as. similar test can be obtained as follows. Let $\beta_n(\zeta; t, \eta)$ be the power of the MP size α test for a simple hypothesis $(0, \zeta)$ against a simple alternative $(\tau t, \eta)$. Then the power $\beta_n(t, \eta)$ of any as. similar test is no greater than $\beta_n(\zeta; t, \eta) + o(\tau^2)$ for any $\zeta \in Z$. Hence, up to $o(\tau^2)$, it is no greater than

$$\bar{\beta}_n(t, \eta) = \inf_\zeta \beta_n(\zeta; t, \eta). \tag{3.5}$$

We do not specify the domain over which inf is taken. It is intuitively clear that the least favorable hypothesis where this infimum is attained lies in a small neighborhood of η.

Let us introduce some notation. We write $l = \log p$ with the same arguments and indices, so that, for example, $l_{\theta,\zeta} = \log p_{\theta,\zeta}$. The differentiation w.r.t. θ and ζ_i is denoted by the superscripts 0 and i, $i = 1, \ldots, k$, e.g.,

$$l^0_{\theta,\zeta} = \frac{\partial}{\partial\theta} l_{\theta,\zeta}, \quad l^i_{\theta,\zeta} = \frac{\partial}{\partial\zeta_i} l_{\theta,\zeta}, \quad l^{001}_{\theta,\zeta} = \frac{\partial}{\partial\theta^2 \partial\zeta_1} l_{\theta,\zeta}, \tag{3.6}$$

etc. We omit the subscript 0 when $\theta = 0$, so that we write $l_\zeta, l^0_\zeta, \ldots$ instead of $l_{0,\zeta}, l^0_{0,\zeta}, \ldots$. We denote by J_ζ and I_ζ the Fisher information (matrix in the latter case) w.r.t. θ and ζ respectively,

$$J_\zeta = E_\zeta(l^0_\zeta)^2, \quad I_\zeta = (I_{\zeta,ij}) \quad \text{with} \quad I_{\zeta,ij} = E_\zeta(l^i_\zeta l^j_\zeta). \tag{3.7}$$

The symmetry of $p_{\theta,\zeta}$ about θ implies that l^0_ζ and l^{0i}_ζ are odd functions, while l^{00}_ζ and l^i_ζ, $i = 1, \ldots, k$, are even. Hence l^0_ζ and l^{0j}_ζ are uncorrelated with l^i_ζ and l^{00}_ζ, $i, j = 1, \ldots, k$,

$$E_\zeta(l^0_\zeta l^i_\zeta) = E_\zeta(l^0_\zeta l^{00}_\zeta) = E_\zeta(l^{0j}_\zeta l^i_\zeta) = E_\zeta(l^{0j}_\zeta l^{00}_\zeta) = 0. \tag{3.8}$$

For simplicity of presentation, we will treat ζ as a univariate parameter, stating only final formulas for the vector case. Differentiation w.r.t. this parameter will be denoted by the superscript 1.

In general, without the symmetry assumption (when l^0_ζ and l^1_ζ are correlated), the least favorable hypothesis to the alternative (3.3) deviates from η by a quantity of order τ (proportionally to the deviation of θ). In the symmetric case this main term vanishes, so that we will seek the minimizer in (3.5) in the form $\zeta = \eta + \tau^2 b$.

The MP test for $(0, \zeta)$ against $(\tau t, \eta)$ is based on the LLR

$$\Lambda_n(\zeta; t, \eta) = \log \frac{p_{n,t,\eta}}{p_{n,0,\zeta}} = \log \prod \frac{p_{\tau t,\eta}(X_i)}{p_{0,\zeta}(X_i)}. \tag{3.9}$$

(We use the notation for the product density similar to (2.4).) It can be written as $\Lambda_n(\zeta; t, \eta) = \Lambda_n(\eta; t, \eta) - \Lambda_n(\eta; 0, \zeta)$ with

$$\Lambda_n(\eta; t, \eta) = \log \frac{p_{n,t,\eta}}{p_{n,0,\eta}}, \quad \Lambda_n(\eta; 0, \zeta) = \log \frac{p_{n,0,\zeta}}{p_{n,0,\eta}}. \tag{3.10}$$

Here $\Lambda_n(\eta; t, \eta)$ is the LLR of distributions differing only by the location parameter with η fixed, hence the formulas in **2.1** are applicable. Using notation (2.8) with obvious modifications, we have by (2.10)

$$\Lambda_n(\eta; t, \eta) = tL^0_{n,\eta} - \tfrac{1}{2}t^2 J_\eta + \tfrac{1}{2}\tau t^2 L^{00}_{n,\eta} + \ldots \tag{3.11}$$

In a similar way we obtain for $\zeta = \eta + \tau^2 b$

$$\Lambda_n(\eta; 0, \eta + \tau^2 b) = \tau b L^1_{n,\eta} - \tfrac{1}{2}\tau^2 b^2 I + \ldots \tag{3.12}$$

Hence

$$\Lambda_n(\eta + \tau^2 b; t, \eta) = tL^0_{n,\eta} - \tfrac{1}{2}t^2 J_\eta + \tau(\tfrac{1}{2}t^2 L^{00}_{n,\eta} - bL^1_{n,\eta}) + \ldots \tag{3.13}$$

This test rejects the hypothesis when

$$\Lambda_n(\eta+\tau^2 b; t, \eta) > c_n(b) \tag{3.14}$$

(suppressing the other arguments like t, η on which c_n depends) with

$$P_{n,0,\eta+\tau^2 b}(\Lambda_n(\eta+\tau^2 b; t, \eta) > c_n(b)) = \alpha \tag{3.15}$$

(or $\alpha + o(\tau^2)$).

Its power is

$$\beta_n(\eta+\tau^2 b; t, \eta) = P_{n,\tau t,\eta+\tau^2 b}(\Lambda_n(\eta+\tau^2 b; t, \eta) > c_n(b)). \tag{3.16}$$

In particular, for $b=0$ this formulas are related to the MP test for $(0,\eta)$ against $(\tau t, \eta)$. We denoted the power of this test by $\beta_n^*(t,\eta)$. We can regard the test (3.14) as an as. efficient test for the same testing problem. Then the difference between the two powers could be directly found by the formula (2.28) if these tests had the same size (up to $o(\tau^2)$). However the $(0,\eta)$-probability of (3.14) to be denoted by α_n' differs from α (which is the $(0,\eta+\tau^2 b)$-probability of the same event) by a quantity of order τ^2. Namely, we have

$$\alpha - \alpha_n' = \tau^2 \frac{\varphi(u_{1-\alpha})}{2t\sqrt{J_\eta}} h + o(\tau^2), \tag{3.17}$$

where

$$h = h(b,t) = b_i a^i - 2b^T I_\eta b, \quad a^i = t^2 E_{0,\eta}(l_\eta^{00} l_\eta^i) + \frac{t u_{1-\alpha}}{\sqrt{J_\eta}} E_{0,\eta}((l_\eta^0)^2 l_0^i). \tag{3.18}$$

This formula is stated for the case of a vector nuisance parameter, meaning the summation over the repeated index i. In this case I_η is the $(k \times k)$ Fisher information matrix.

The formula (3.17–18) will be derived in **5.1**. Using this formula we obtain here an asymptotic formula for $\beta_n^*(t,\eta) - \bar\beta_n(t,\eta)$ (see (3.5)).

Denote by $\beta_n^{*'}(t,\eta)$ the power of the MP test for $(0,\eta)$ against $(\tau t,\eta)$ (based on $\Lambda_n(\eta; t, \eta)$) of size α_n'. Then this test has the same size as the test (3.14), hence we can apply the formula (2.28). Comparing (3.11) with (3.13) we see that Δ_n as in (2.21) equals $\tau b L^1_{n,\eta}$, so that denoting by (L_η^0, L_η^1) a bivariate normal r.v. to which $(L^0_{n,\eta}, L^1_{n,\eta})$ converges in distribution under $P_{0,\eta}$ we see that (2.27) holds with

$$\Pi = bL_\eta^1, \quad \Lambda = tL_\eta^0 - \frac{1}{2}t^2 J_\eta. \tag{3.19}$$

Due to (3.8) L_η^0 and L_η^1 are independent, so that the conditional variance in (2.28) equals the unconditional one, which is $b^T I_\eta b$. Thus by (2.28) and (2.17)

$$\beta_n^{*'}(t,\eta) - \beta_n(\eta+\tau^2 b; t,\eta) = \frac{1}{2}\tau^2 \frac{1}{t\sqrt{J}} \varphi(u_{1-\alpha} - t\sqrt{J}_\eta) b^T I_\eta b. \tag{3.20}$$

Comparing (3.17–18) with (2.18) we see that the tests based on $\Lambda_n(\eta; t, \eta)$ of sizes α and α' satisfy (2.18) with $\delta_n = \frac{1}{2}\tau^2 h$. Hence by (2.19)

$$\beta_n^*(t,\eta) - \beta_n^{*'}(t,\eta) = \frac{1}{2}\tau^2 \frac{1}{t\sqrt{J_\eta}} \varphi(u_{1-\alpha} - t\sqrt{J_\eta}) h + o(\tau^2), \tag{3.21}$$

where h is given by (3.18).

Thus we obtain from (3.20) and (3.21)

$$\beta_n^*(t,\eta) - \beta_n(\eta+\tau^2 b; t,\eta) = \frac{1}{2}\tau^2 \frac{1}{t\sqrt{J_\eta}}\varphi(u_{1-\alpha} - t\sqrt{J_\eta})(h + b^T I_\eta b) + o(\tau^2). \tag{3.22}$$

It is seen from (3.18) that $h(b)+b^T I_\eta b$ is a quadratic function of b. It attains its maximum at

$$b_0 = \frac{1}{2} I_\eta^{-1} a, \tag{3.23}$$

where $a = a(t)$ is given in (3.18), and its maximal value is $\frac{1}{4}a^T I_\eta^{-1} a$. Since β_n^* in (3.22) does not depend on b, maximization of (3.22) corresponds to minimization of $\beta_n(\eta+\tau^2 b; t,\eta)$. Therefore (see (3.5))

$$\beta_n^*(t,\eta) - \bar{\beta}_n(t,\eta) = \frac{1}{8}\tau^2 \frac{1}{t\sqrt{J}}\varphi(u_{1-\alpha} - t\sqrt{J_\eta}) a^t I_\eta^{-1} a + o(\tau^2). \tag{3.24}$$

Recall that the power $\beta_n(t,\eta)$ of any as. similar test is no greater than $\bar{\beta}_n(t,\eta)$ (see the argument before (3.5)). Hence the RHS of (3.24) provides a lower bound for the power loss $\beta_n^*(t,\eta) - \beta_n(t,\eta)$ of any as. similar test.

Remark. It is seen from (3.23), (3.18) that $h(b_0,t) = 0$, i.e., $\alpha' = \alpha + o(\tau^2)$ for $b = b_0$, where α' is the size of the test based on $\Lambda_n(\eta+\tau^2 b_0; t,\eta)$ (see (3.14), (3.17)). One can check that this test is as. similar (i.e., of size $\alpha + o(\tau^2)$) on any neighborhood $(\eta - C\tau^2, \eta + C\tau^2)$ of η shrinking at a rate of τ^2.

For convenience of comparison with the bound given in the next section, restate (3.25) for the alternative $(\tau t/\sqrt{J_\eta}, \eta)$ rather than $(\tau t, \eta)$ (this normalization will be essential for derivation of that bound). Denote $d = (d^1, \ldots, d^k)$ with

$$d^i = d^i(t,\eta) = t E_{0,\eta}(l_\eta^{00} l_\eta^i) + u_{1-\alpha} E_{0,\eta}((l_\eta^0)^2 l_0^i). \tag{3.25}$$

Then for the power $\beta_n(t,\eta)$ of an arbitrary as. similar test we have

$$\beta_n^*(t/\sqrt{J_\eta}, \eta) - \beta_n(t/\sqrt{J_\eta}, \eta) \geq \tau^2 B_1(t,\eta) + o(\tau^2), \tag{3.26}$$

where

$$B_1(t,\eta) = \frac{t}{8 J_\eta^2}\varphi(u_{1-\alpha} - t) d^T I_\eta^{-1} d. \tag{3.27}$$

3.2 A test attaining the local bound

Here we demonstrate an as. similar test on which the above bound is attained at a given alternative $(\tau t, \eta)$. Note that this test depends on the chosen alternative and is not even first-order efficient against alternatives with $\zeta \neq \eta$. For simplicity of presentation we treat only a univariate parameter ζ.

In the previous subsection we derived the lower bound for the power loss as the power loss of a specific test, namely, the one based on the LLR $\Lambda_n(\eta+\tau^2 b_0; t,\eta)$. As pointed out in Remark 3.1, this test retains the size $\alpha + o(\tau^2)$ on small neighborhoods of η shrinking at a rate of τ^2. We will show that this test can be modified to become as. similar (i.e., of size $\alpha + o(\tau^2)$ uniformly on compact subsets of Z) retaining the power loss at $(\tau t, \eta)$.

Consider the statistic

$$S_n = S_n(t, \eta) = L^0_{n,\eta} + \frac{1}{2}\tau\Big(tL^{00}_{n,\eta} - \frac{a}{tI_\eta}L^1_{n,\eta}\Big). \tag{3.28}$$

It differs from $\Lambda_n(\eta + \tau^2 b_0; t, \eta)$ (see (3.13), (3.17), and (3.23)) by dropping the terms of order τ^2 and by additive and multiplicative constants. As we pointed out, the power loss is determined by the terms of order τ in the stochastic expansion of the test statistic, so that the test based on S_n has the same power loss under $(\tau t, \eta)$ as the test based on $\Lambda_n(\eta + \tau^2 b_0; t, \eta)$.

Denote $q(\zeta) = \big(E_\zeta(l^0_\eta)^2\big)^{-1/2}$. Let $\hat\zeta_n$ be the maximum likelihood estimate (MLE) for ζ. As the first step, we studentize S_n, i.e., we consider

$$S_{n1} = S_n q(\hat\zeta_n). \tag{3.29}$$

It is well known that under $P_{n,0,\zeta}$

$$\hat\zeta_n = \zeta + \tau I_\zeta^{-1} L^1_{n,\zeta} + \dots . \tag{3.30}$$

Putting this into (3.29) yields

$$S_{n1} = S_n(q(\zeta) + \tau q'(\zeta) I_\zeta^{-1} L^1_{n,\zeta}) + \dots \tag{3.31}$$

Next define

$$S_{n2} = S_{n1} - \tau S_n q'(\eta) I_\eta^{-1} L^1_{n,\eta}. \tag{3.32}$$

It is seen from (3.31) and (3.32) that S_{n2} under $P_{n,0,\eta}$ differs from $S_n q(\eta)$ by terms of order τ^2. We will construct a statistic S_{n3} differing from S_{n2} by terms of order τ^2 of its stochastic expansion, which determines an as. similar test. Hence the test based on S_{n3} is as. similar and the statistic $S_{n3}/q(\eta)$ coincides under $P_{n,0,\eta}$ with S_n as in (3.28) up to (including) terms of order τ. Therefore this test has the same power loss for the alternative $(\tau t, \eta)$ as the one based on S_n, thus attaining the lower bound for this alternative. Thus the test based on S_{n3} will have the desired properties.

To construct the required correction of S_{n2}, consider its stochastic expansion. We have from (3.28), (3.31), and (3.32) that under $P_{n,0,\zeta}$

$$S_{n2} = q(\zeta)L^0_{n,\eta} + \frac{1}{2}\tau q(\zeta)\Big(tL^{00}_{n,\eta} - \frac{a}{tI_\eta}L^1_{n,\eta}\Big) - \tau q'(\eta) I_\eta^{-1} L^1_{n,\eta} L^0_{n,\eta} +$$

$$+\tau q'(\zeta) I_\zeta^{-1} L^1_{n,\zeta} L^0_{n,\eta} + \dots \tag{3.33}$$

The key argument is that the τ-term in the corresponding E-expansion vanishes, so that the E-expansion has the form

$$P_{n,0,\zeta}(S_{n2} < x) = \Phi(x) + \tau^2 Q(x, \zeta)\phi(x) + o(\tau^2). \tag{3.34}$$

Indeed, since l^0_η is odd and l^{00}_η and l^1_η are even, the third moment of l^0_η vanishes and $L^0_{n,\eta}$ is uncorrelated with $L^{00}_{n,\eta}$ and $L^1_{n,\eta}$ under $P_{n,0,\zeta}$ for any $\zeta \in Z$. Hence our claim follows from the form of the one-term E-expansion given in **5.2**. Now (3.34) can be rewritten as

$$P_{n,0,\zeta}(S_{n2} - \tau^2 Q(x, \zeta) < x) = \Phi(x) + o(\tau^2). \tag{3.35}$$

Then $S_{n3} = S_{n2} - \tau^2 Q(x, \hat\zeta_n)$ has the required property.

4 Local asymptotic minimaxity in terms of the power loss

4.1 The lower bound

The bound given in the previous section can be attained by some tests, which, however, depend on the chosen η. This resembles the superefficiency effect in estimation, where a lower risk than the regular (Cramér–Rao) bound can be attained in a vicinity of a given parameter point by estimators tuned to this point at the expense of increase of the risk at some other points. Like in estimation, this suggests the minimax approach where a test is characterized by the maximal loss over a neighborhood in the parameter space. Hence, for a fixed $t > 0$, we characterize an as. similar test with power $\beta_n(t, \zeta)$ at the alternative $(t/\sqrt{J_\eta}, \eta)$ by

$$r_n(t, K) = \sup_K (\beta_n^*(t/\sqrt{J_\zeta}, \zeta) - \beta_n(t/\sqrt{J_\zeta}, \zeta)), \tag{4.1}$$

where $K \subset Z$ is a (small) neighborhood of η and $\beta_n^*(t, \zeta)$, as before, is the power of the MP test for $(0, \zeta)$ against $(\tau t, \zeta)$. For a univariate ζ we take K to be a finite interval. We normalize t by $\sqrt{J_\zeta}$ in order to exclude the effect of variation of J_ζ on the power loss when taking the supremum. Later on we will point out where this normalization comes into effect techically. For notational convenience, denote $t(\zeta) = t/\sqrt{J_\zeta}$.

Let $\pi(d\zeta)$ be the uniform distribution on K (though many arguments to follow remain valid for more general π). Then, obviously,

$$r_n(t, K) \geq \int_K (\beta_n^*(t(\zeta), \zeta) - \beta_n(t(\zeta), \zeta))\pi(d\zeta). \tag{4.2}$$

Denote by $\mathcal{A}(\alpha)$ the class of tests with average size α, i.e., tests such that

$$\int_K \beta_n(0, \zeta)\pi(d\zeta) = \alpha. \tag{4.3}$$

Let $\bar{\beta}_n(t, \zeta)$ be the power of the test in $\mathcal{A}(\alpha)$ maximizing

$$\beta_n^\pi(t) = \int_K \beta_n(t(\zeta), \zeta)\pi(d\zeta). \tag{4.4}$$

Denote this maximal average power by $\bar{\beta}_n^\pi(t)$. The size of any as. similar test satisfies (4.3) with $\alpha + o(\tau^2)$ in the RHS. Hence its average power is no greater than $\bar{\beta}_n^\pi(t) + o(\tau^2)$. Thus for $r_n(t, K)$ related to an arbitrary as. similar test we obtain by (4.2), (4.4) the lower bound

$$r_n(t, K) \geq \int_K \beta_n^*(t(\zeta), \zeta)\pi(d\zeta) - \bar{\beta}_n^\pi(t) + o(\tau^2). \tag{4.5}$$

Denote by $P_{n,0,\zeta}$ and $P_{n,1,\zeta}$ the probability measures with densities

$$\begin{aligned} p_{n,0,\zeta}(\mathbf{x}) &= \prod p_{0,\zeta}(x_i), \\ p_{n,1,\zeta}(\mathbf{x}) &= \prod p_{\tau t(\zeta),\zeta}(x_i), \quad \mathbf{x} = (x_1, \ldots, x_n). \end{aligned}$$

(We keep t fixed and suppress it in the notation of the alternative densities and distributions.) Let

$$p^\pi_{n,1}(\mathbf{x}) = \int_K p_{n,1,\zeta}(\mathbf{x})\pi(d\zeta), \quad p^\pi_{n,0}(\mathbf{x}) = \int_K p_{n,0,\zeta}(\mathbf{x})\pi(d\zeta). \tag{4.6}$$

Denote by $P^\pi_{n,1}, P^\pi_{n,0}, E^\pi_{n,1}, E^\pi_{n,0}$ the corresponding distributions and expectations. It is seen from (4.3), (4.4) that the test maximizing $\beta^\pi_n(t)$ over $\mathcal{A}(\alpha)$ is the MP size α test for the simple hypothesis $P^\pi_{n,0}$ against the simple alternative $P^\pi_{n,1}$. This Bayes test is based on the LLR

$$\Lambda^\pi_n(t) = \log\frac{p^\pi_{n,1}(\mathbf{X})}{p^\pi_{n,0}(\mathbf{X})}, \quad \mathbf{X} = (X_1, \ldots, X_n). \tag{4.7}$$

While this test has size α w.r.t. $P^\pi_{n,0}$, its size $\bar\beta_n(0,\zeta)$ w.r.t. $P_{n,0,\zeta}$ differs from α by a quantity of order τ^2. We will show that $\Lambda^\pi_n(t)$ can be amended in terms of order τ^2 so that the average power $\bar\beta^\pi_n(t)$ changes only by $o(\tau^2)$ and the corresponding test is as. similar. In other words, we will construct an as. similar test with power $\tilde\beta_n(t,\zeta)$ such that the corresponding average power $\tilde\beta^\pi_n(t)$ equals $\bar\beta^\pi_n(t)$ up to $o(\tau^2)$,

$$\tilde\beta^\pi_n(t) - \bar\beta^\pi_n(t) = o(\tau^2). \tag{4.8}$$

Therefore the lower bound (4.5) can be restated as

$$r_n(t,K) \geq \int_K (\beta^*_n(t(\zeta),\zeta) - \tilde\beta_n(t(\zeta),\zeta))\pi(d\zeta) + o(\tau^2). \tag{4.9}$$

Then the difference in the integrand can be evaluated by formula (2.28).

In this way we will obtain that

$$\beta^*_n(t(\zeta),\zeta) - \tilde\beta_n(t(\zeta),\zeta) = \tau^2 B^*(t,\zeta) + o(\tau^2),$$

with B^* given by (4.31).

Like in the estimation theory, to obtain a lower bound for the power loss at a given point η we pass to a limit as $n \to \infty$ in the bound (4.9) for $nr_n(t,K)$ with K taken to be an interval containing η as an interior point, and then we pass to a limit as K shrinks to η. Under appropriate regularity conditions $B^*(t,\zeta)$ is continuous in ζ, so that the average value of B^* converges to $B^*(t,\eta)$ when K shrinks to η. Thus, denoting by $\mathcal{S}$ the class of as. similar tests and by $\beta^\phi_n(t,\zeta)$ the power of the test $\phi \in \mathcal{S}$ we obtain

$$\lim_{K\downarrow\eta}\liminf_{n\to\infty}\inf_{\phi\in\mathcal{S}}\sup_{\zeta\in K} n(\beta^*_n(t/\sqrt{J_\zeta},\zeta) - \beta^\phi_n(t/\sqrt{J_\zeta},\zeta)) \geq B^*(t,\eta). \tag{4.10}$$

To carry out the program outlined above, we derive a stochastic expansion for $\Lambda^\pi_n(t)$ given by (4.7). We assume that the "true value" of the nuisance parameter is some η, interior to K. Write $p^\pi_{n,1}(\mathbf{X})$ (see (4.6)) as

$$p^\pi_{n,1}(\mathbf{X}) = p_{n,1,\eta}(\mathbf{X})\, q^\pi_{n,1}(\mathbf{X}),$$

where

$$q^\pi_{n,1}(\mathbf{X}) = \int_K \exp\left(\sum \log\frac{p_{\tau t(\zeta),\zeta}(X_i)}{p_{\tau t(\eta),\eta}(X_i)}\right)\pi(d\zeta). \tag{4.11}$$

Similarly, we write

$$p^{\pi}_{n,0}(\mathbf{X}) = p_{n,0,\eta}(\mathbf{X})q^{\pi}_{n,0}(\mathbf{X}),$$

where $q^{\pi}_{n,0}(\mathbf{X})$ is defined by (4.11) with $t = 0$. Thus the LLR (4.7) can be written as

$$\Lambda^{\pi}_{n}(t) = \log\frac{p_{n,1,\eta}(\mathbf{X})}{p_{n,0,\eta}(\mathbf{X})} + \log\frac{q^{\pi}_{n,1}(\mathbf{X})}{q^{\pi}_{n,0}(\mathbf{X})}. \tag{4.12}$$

Note that the first term in the RHS of (4.12) is the LLR of the distributions with densities $p_{n,1,\eta}$ and $p_{n,0,\eta}$, which determines the MP test for $(0,\eta)$ against $\tau t(\eta),\eta$. This is the test which we would apply if η were known and relative to which we calculate the power loss.

We will return to this test later on, and now we consider the second term in the RHS of (4.12). Denote the exponent in (4.11) by $M_n(t,\zeta,\eta)$,

$$M_n(t,\zeta,\eta) = \sum(l_{\tau t(\zeta),\zeta}(X_i) - l_{\tau t(\eta),\eta}(X_i)).$$

The LLR M_n has a nontrivial limit when the deviation of ζ from η is of order τ. Hence we make the substitution $\zeta - \eta = \tau z$. Then the domain of integration w.r.t. z extends at a rate of $\sqrt{n}$, and the tails of the integrand decrease sufficiently fast, so that with high accuracy the integration can be extended to the whole real line.

The difference $t(\zeta) - t(\eta)$ for $\zeta = \eta + \tau z$ is

$$\tau t/\sqrt{J_\zeta} - \tau t/\sqrt{J_\eta} = -\frac{1}{2}\tau^2 tz J_\eta^{-3/2} J^1_\eta + o(\tau^2).$$

(As before, the superscript 1 of J_η means the derivative w.r.t. η.) By the Taylor series expansion around η we obtain

$$M_n(t,\eta+\tau z,\eta) = -\frac{1}{2}\tau^2 tz\frac{J^1_\eta}{J^{3/2}_\eta}\sum l^0_{\tau t(\eta),\eta}$$
$$+\tau z\sum l^1_{\tau t(\eta),\eta} + \frac{1}{2}(\tau z)^2\sum l^{11}_{\tau t(\eta),\eta} + \frac{1}{6}(\tau z)^3\sum l^{111}_{\tau t(\eta),\eta} + \dots$$

Next we use the Taylor expansions

$$\begin{aligned}
l^0_{\tau t(\eta),\eta} &= l^0_{0,\eta} + \tau t(\eta) l^{00}_{0,\eta} + \dots,\\
l^1_{\tau t(\eta),\eta} &= l^1_{0,\eta} + \tau t(\eta) l^{01}_{0,\eta} + \tau^2 t^2(\eta) l^{001}_{0,\eta} + \dots,\\
l^{11}_{\tau t(\eta),\eta} &= l^{11}_{0,\eta} + \tau t(\eta) l^{011}_{0,\eta} + \dots,\\
l^{111}_{\tau t(\eta),\eta} &= l^{111}_{0,\eta} + \dots
\end{aligned}$$

Similarly to (2.8) we denote by $L^0_{n,\eta}, L^{00}_{n,\eta},\dots$ the centered and normalized sums of the corresponding derivatives. Note that

$$E_{0,\eta}l^1_{0,\eta} = E_{0,\eta}l^{01}_{0,\eta} = E_{0,\eta}l^{011}_{0,\eta} = 0, \quad E_{0,\eta}l^{00}_{0,\eta} = -J_\eta, \quad E_{0,\eta}l^{11}_{0,\eta} = -I_\eta. \tag{4.13}$$

Denote $m_{001} = E_{0,\eta}l^{001}_{0,\eta}$. Then

$$\begin{aligned}
M_n(t,\eta+\tau z,\eta) = & -\frac{1}{2}\tau zt J_\eta^{-3/2} J^1_\eta L^0_{n,\eta} + \frac{1}{2}\tau z t^2\frac{J^1_\eta}{J_\eta}\\
& + zL^1_{n,\eta} + \tau\frac{zt}{\sqrt{J_\eta}}L^{01}_{n,\eta} + \frac{1}{2}\tau z\frac{t^2}{J_\eta}m_{001}\\
& - \frac{1}{2}z^2 I_\eta + \frac{1}{2}\tau z^2 L^{11}_{n,\eta} + \frac{1}{6}\tau z^3 m_{111} + \dots
\end{aligned}$$

Thus the integrand in (4.11) is

$$\exp\left(-\tfrac{1}{2}z^2 I_\eta + zL^1_{n,\eta}\right)\left[1+\tau(zU_{n,\eta}(t)+\tfrac{1}{2}z^2L^{11}_{n,\eta}+\tfrac{1}{6}z^3 m_{111})\right]+\dots, \tag{4.14}$$

where

$$U_{n,\eta}(t) = -\frac{t}{2}\frac{J^1_\eta}{J^{3/2}_\eta}L^0_{n,\eta} + \frac{1}{2}t^2\frac{J^1_\eta}{J_\eta} + \frac{t}{\sqrt{J_\eta}}L^{01}_{n,\eta} + \frac{1}{2}\frac{t^2}{J_\eta}m_{001}. \tag{4.15}$$

Integrating w.r.t. $\pi(d\zeta)$ we obtain the ratio $q^\pi_{n,1}(x)/q^\pi_{n,0}(x)$ in the form

$$\frac{q^\pi_{n,1}(x)}{q^\pi_{n,0}(x)} = \frac{A+\tau A_1}{A+\tau A_0} = 1+\tau\frac{A_1-A_0}{A}\dots, \tag{4.16}$$

where A and A_1 are the integrals of the main term and the order τ term in (4.14) and A_0 is obtained from A_1 by putting $t=0$. The terms in (4.14) not containing t cancel when taking the difference $A_1 - A_0$ and the calculation of $(A_1-A_0)/A$ reduces to writing down the mean value of the corresponding normal distribution, which yields

$$\frac{q^\pi_{n,1}(x)}{q^\pi_{n,0}(x)} = 1+\tau\frac{L^1_{n,\eta}}{I_\eta}U_{n,\eta}(t)+\dots \tag{4.17}$$

The logarithm of this ratio to be used in (4.12) is just the τ-term in (4.17) (up to terms of higher order).

Denote the first term in the RHS of (4.12) by

$$\Lambda_n(t(\eta),\eta) = \log\frac{p_{n,1,\eta}(\mathbf{X})}{p_{n,0,\eta}(\mathbf{X})}. \tag{4.18}$$

As we pointed out, this is the LLR of the two distributions for η given, with which we compare our test statistics. In view of (4.17) we can rewrite (4.12) as

$$\Lambda^\pi_n(t) = \Lambda_n(t(\eta),\eta)+\tau\frac{L^1_{n,\eta}}{I_\eta}U_{n,\eta}(t)+\dots \tag{4.19}$$

This relation could be used for the application of the formula (2.28) (see (2.21), (2.27)), if the test based on $\Lambda^\pi_n(t)$ were as. similar. We will show that it can be made as. similar by a correction in terms of order τ^2 which do not affect its power.

For that we need a stochastic expansion for $\Lambda^\pi_n(t)$, which is obtained by writing down that for $\Lambda_n(t(\eta),\eta)$ in (4.19). The latter LLR can be treated as in the case of a univariate parameter. Hence by (2.10) it is

$$\Lambda_n(t(\eta),\eta) = \frac{t}{\sqrt{J_\eta}}L^0_{n,\eta} - \frac{1}{2}t^2 + \frac{1}{2}\tau\frac{t^2}{J_\eta}L^{00}_{n,\eta}+\dots$$

Thus by (4.7), (4.12), and (4.17)

$$\Lambda^\pi_n(t) = \frac{t}{\sqrt{J_\eta}}L^0_{n,\eta} - \frac{1}{2}t^2 + \tau V_{n,\eta}(t)+\dots, \tag{4.20}$$

where, with $U_{n,\eta}(t)$ given by (4.15),

$$V_{n,\eta}(t) = \frac{t^2}{2J_\eta} L^{00}_{n,\eta} + \frac{L^1_{n,\eta}}{I_\eta} U_{n,\eta}(t). \tag{4.21}$$

We denote the power of the Bayes test based on $\Lambda^\pi_n(t)$ by $\bar\beta^\pi_n(t)$,

$$\bar\beta^\pi_n(t) = P^\pi_{n,1}(\Lambda^\pi_n(t) > c_n),$$

where c_n is such that this test is of size α w.r.t. $P^\pi_{n,0}$,

$$P^\pi_{n,0}(\Lambda^\pi_n(t) > c_n) = \int_K P_{n,0,\zeta}(\Lambda^\pi_n(t) > c_n)\pi(d\zeta) = \alpha. \tag{4.22}$$

Now we will construct an as. similar test of the form $S_{n,t} > c$ with power $\tilde\beta_n(t,\zeta)$ whose test statistic differs from $\Lambda^\pi_n(t)$ by terms of order τ^2. Since $\Lambda^\pi_n(t)$ is the LLR statistic in testing $P^\pi_{n,0}$ vs $P^\pi_{n,1}$, by the formula (2.28) the power

$$\tilde\beta^\pi_n(t) = \int_K \tilde\beta_n(t,\zeta)\pi(d\zeta) \tag{4.23}$$

satisfies (4.8).

To this end, consider the E-expansion for the distribution of $\Lambda^\pi_n(t)$ under $P_{n,0,\zeta}$. The main term is the normal distribution $N(-\frac{1}{2}t^2, t^2)$. Denote its d.f. and density by $F_{0,t}(x)$ and $f_{0,t}(x)$. (It is important that the main term $F_{0,t}$ of the E-expansion does not depend on ζ, which is due to the normalization of t by $\sqrt{J_\zeta}$.) The term of order τ in this E-expansion vanishes because it consists of the term with the 3rd moment of $l_{0,\zeta}$ vanishing since this is an even function and the conditional expectation

$$E\left(V_\zeta(t)\,\middle|\, tJ_\zeta^{-1/2}L^0_\zeta - \tfrac{1}{2}t^2 = c_n\right), \tag{4.24}$$

where $V_\zeta(t)$ depends on a zero-mean normal vector

$$(L^0_\zeta, L^1_\zeta, L^{00}_\zeta, L^{01}_\zeta) \tag{4.25}$$

to which $(L^0_{n,\zeta}, L^1_{n,\zeta}, L^{00}_{n,\zeta}, L^{01}_{n,\zeta})$ converges in distribution in the same way as $V_{n,\zeta}(t)$ depends on $(L^0_{n,\zeta},\ldots)$ (see (4.15), (4.21)). (The formula for the one-term E-expansion of a statistic like (4.20) is given in **5.2**.) Due to the symmetry properties (see (3.8)), L^0_ζ is independent of L^{00}_ζ and L^1_ζ is independent of $(L^0_\zeta, L^{01}_\zeta)$. Using these relations it is readily verified that the conditional expectation (4.24) vanishes.

Denote by c the $(1-\alpha)$-quantile of $F_{0,t}$, viz., $c = tu_{1-\alpha} - \frac{1}{2}t^2$. Then

$$P_{n,0,\zeta}\left(\Lambda^\pi_n(t) > c\right) = 1 - F_{0,t}(c) - \tau^2 f_{0,t}(c)g_\zeta(c) + o(\tau^2), \tag{4.26}$$

where $1 - F_{0,t}(c) = \alpha$ and $g_\zeta(c)$ is a certain polynomial in the E-expansion evaluated at $x = c$.

The particular form of $g_\zeta(c)$ is immaterial, we only need that it is a sufficiently smooth function of ζ (under certain regularity conditions). It is seen from (4.26) that

$$P_{n,0,\zeta}\left(\Lambda^\pi_n(t) + \tau^2 g(\zeta) > c\right) = \alpha + o(\tau^2). \tag{4.27}$$

Obviously, this relation will continue to hold if we replace $g(\zeta)$ by $g(\hat{\zeta}_n)$, where $\hat{\zeta}_n$ is any consistent estimator of ζ. Hence setting

$$S_{n,t} = \Lambda_n^\pi(t) + \tau^2 g(\hat{\zeta}_n) \tag{4.28}$$

we have

$$\tilde{\beta}_n(0,\zeta) = P_{n,0,\zeta}(S_{n,t} > c) = \alpha + o(\tau^2)$$

for any ζ interior to K, so that this test is as. similar. As we pointed out above, the averaged power (4.23) for

$$\tilde{\beta}_n(t,\zeta) = P_{n,1,\zeta}(S_{n,t} > c)$$

differs from $\bar{\beta}_n(t,\zeta)$ by $o(\tau^2)$, i.e., it satisfies (4.8).

Now we obtain an asymptotic formula for the integrand in (4.9), $\beta_n^*(t,\zeta) - \tilde{\beta}_n(t,\zeta)$. Recall that $\beta_n^*(t,\zeta)$ is the power of the MP size α test for $P_{n,0,\zeta}$ vs $P_{n,1,\zeta}$, which is based on $\Lambda_n(t(\zeta),\zeta)$ given by (4.18), and $\tilde{\beta}_n(t,\zeta)$ is the power in this testing problem of the test based on the statistic (4.28) having the same size up to $o(\tau^2)$. Hence by (4.19)

$$\sqrt{n}\,(\Lambda_n(t(\zeta),\zeta) - S_{n,t}) = -\frac{L^1_{n,\zeta}}{I_\zeta} U_{n,\zeta}(t) + \ldots, \tag{4.29}$$

where $U_{n,\zeta}(t)$ is given by (4.15). Therefore (2.27) is fulfilled with

$$\Pi = \frac{L^1_\zeta}{I_\zeta} U_\zeta(t), \quad \Lambda = \frac{t}{\sqrt{J_\zeta}} L^0_\zeta - \frac{1}{2}t^2,$$

where $U_\zeta(t)$ is given by (4.15) with $L^0_{n,\zeta}$ and $L^{01}_{n,\zeta}$ replaced by the corresponding components of the limiting vector (4.25). In our case $e^c f_0(c)$ in (2.28) is $e^c f_{0,t}(c)$ with $f_{0,t}(c)$ the density of $N(-\frac{1}{2}t^2, t^2)$ and c the $(1-\alpha)$-quantile of this distribution, so that $e^c f_{0,t}(c) = t^{-1}\varphi(u_{1-\alpha} - t)$ (cf. (2.17)). The condition $\Lambda = c$ becomes $L^0_\zeta = u_{1-\alpha}\sqrt{J_\zeta}$. Thus by (2.28)

$$\beta_n^*(t,\zeta) - \tilde{\beta}_n(t,\zeta) = \tau^2 B^*(t,\zeta) + o(\tau^2), \tag{4.30}$$

where

$$B^*(t,\zeta) = \frac{1}{2t}\varphi(u_{1-\alpha} - t)\,\mathrm{Var}\left[L^1_\zeta I_\zeta^{-1} U_\zeta(t) \mid L^0_\zeta = u_{1-\alpha}\sqrt{J_\zeta}\right].$$

By a routine computation of the conditional variance we obtain for $\zeta \in R^k$

$$B^*(t,\zeta) = \frac{t}{8J_\zeta^2}\varphi(u_{1-\alpha} - t)\left[d^T I_\zeta^{-1} d + 4I_\zeta^{ij}(J_\zeta E(l_\zeta^{0i} l_\zeta^{0j}) - E(l_\zeta^0 l_\zeta^{0i})E(l_\zeta^0 l_\zeta^{0j}))\right], \tag{4.31}$$

where $d = d(t,\zeta)$ is given by (3.25), $I_\zeta^{-1} = \left(I_\zeta^{ij}\right)$, and E stands for $E_{0,\zeta}$.

By putting (4.30) into (4.9) we arrive at (4.10).

It is seen that B^* given by (4.31) differs from B_1 given by (3.27) by the term of the form

$$I^{ij}(JE(l^{0i}l^{0j}) - E(l^0 l^{0i})E(l^0 l^{0j})) \geq 0$$

by a version of the Cauchy–Bunyakovsky inequality. Hence $B^* \geq B_1$, i.e., (4.10) provides a more accurate lower bound than the one in Section 3. In the next section we indicate tests for which it is attained.

4.2 Locally asymptotically minimax tests

The bound (4.10) was obtained as the power loss of the test based on the statistic $S_{n,t}$ given by (4.28). This test, however, depends on the chosen point ζ, see (4.20), (4.21), and hence attains the LAM bound (4.10) only at this point ζ. Here we construct an as. similar test which attains the LAM bound for any $\zeta \in Z$.

Note that both the bound and the test depend on the (local) alternative in terms of the parameter of interest specified as $\theta = t/\sqrt{J_\zeta n}$, $t > 0$. The alternatives $(t/\sqrt{J_\zeta n}, \zeta)$, $\zeta \in Z$ form the "level surface" of asymptotically equal power in the parameter space.

To construct the required test, we start with the statistic (cf. (4.20))

$$S_n(\zeta) = \frac{t}{\sqrt{J_\zeta}} L^0_{n,\zeta} + \tau \bar{V}_{n,\zeta}(t). \tag{4.32}$$

Up to the nonrandom term $\frac{1}{2}t^2$ this statistic has the same stochastic expansion to within the terms of order τ as $\Lambda_n(t(\zeta), \zeta)$ in (4.20) with $V_{n,\zeta}(t)$ substituted by $\bar{V}_{n,\zeta}(t)$ given by (4.21), where $U_{n,\zeta}(t)$ is to be replaced by

$$\bar{U}_{n,\zeta}(t) = \frac{1}{2}t^2 \frac{J^1_\zeta}{J_\zeta} + \frac{1}{2}\frac{t^2}{J_\zeta} m_{001} = U_{n,\zeta}(t) - \Big(-\frac{t}{2}\frac{J^1_\zeta}{J^{3/2}_\zeta} L^0_{n,\zeta} + \frac{t}{\sqrt{J_\zeta}} L^{01}_{n,\zeta}\Big) \tag{4.33}$$

(cf. (4.15)). Now we substitute the MLE $\hat{\zeta}_n$ for ζ in $S_n(\zeta)$. Using the stochastic expansion for $\hat{\zeta}_n$ as in (3.30) and applying the Taylor formula to $S_n(\hat{\zeta}_n)$ we obtain the leading term as in (4.20) and the two terms of order τ which were dropped when replacing $U_{n,\zeta}(t)$ by $\bar{U}_{n,\zeta}(t)$ (see (2)). Together with $\tau\bar{V}_{n,\hat{\zeta}_n}(t) = \tau\bar{V}_{n,\zeta}(t) + \tau^2(\ldots)$ they constitute the stochastic expansion to within terms of order τ as in (4.20) (up to the dropped constant $\frac{1}{2}t^2$), which we aimed at. By an argument similar to the construction of $S_{n,t}$ (see (4.24)–(4.28)) we can correct $S_n(\hat{\zeta}_n)$ in terms of order τ^2 to obtain an as. similar test.

5 Appendix

5.1 Proof of (3.17–18)

Here we outline the proof of the formulas (3.17–18) for the difference $\alpha - \alpha'$, where

$$\begin{aligned} \alpha &= P_{n,0,\eta+\tau^2 b}(\Lambda_n(\eta + \tau^2 b; t, \eta) > c_n(b)), \\ \alpha' &= P_{n,0,\eta}(\Lambda_n(\eta + \tau^2 b; t, \eta) > c_n(b)) \end{aligned}$$

(see (3.15) and the definition before (3.17)). Since η and b are now fixed, we will often suppress the corresponding indices. In particular, we will suppress the subscript η of L^0_n, J, etc. For simplicity we write $P_{n,0} = P_{n,0,\eta}$ and $P_{n,b} = P_{n,0,\eta+\tau^2 b}$. By (3.13), the event $\Lambda_n(\eta + \tau^2 b; t, \eta) > c_n(b)$ is equivalent to $T_n > a_n$, where

$$T_n = \Lambda_n(\eta + \tau^2 b; t, \eta) + \tfrac{1}{2}t^2 J = tL^0_n + \tau(\tfrac{1}{2}t^2 L^{00}_n - bL^1_n) + \ldots \tag{5.1}$$

and a_n differs from $c_n(b)$ by $\frac{1}{2}t^2 J$. The effect of the omitted terms in (5.1) on the difference $\alpha - \alpha'$ is $o(\tau^2)$, so that they can be neglected (though their effect on each of the probabilities

α and α' is of order τ^2). Denote for brevity $\Lambda_n = \Lambda_n(\eta+\tau^2 b; t, \eta)$ (see (3.10), (3.12)). By (3.10)

$$\frac{dP_{n,b}}{dP_{n,0}} = \exp(\Lambda_n).$$

Then

$$\alpha - \alpha' = (P_{n,b} - P_{n,0})(T_n > a_n) = E_{n,0}\mathbf{1}\{T_n > a_n\}\left(e^{\Lambda_n} - 1\right).$$

By (3.12) $e^{\Lambda_n} - 1 \simeq \tau b L_n^1 + \frac{1}{2}\tau^2 b^2\left((L_n^1)^2 - I\right)$. Let $q_n(x^0, x^1, x^{00})$ denote the joint density of (L_n^0, L_n^1, L_n^{00}). Then

$$\alpha - \alpha' = \iiint\limits_{\{T_n > a_n\}} (\tau b x^1 + \frac{1}{2}\tau^2 b^2((x^1)^2 - I)) q_n(x^0, x^1, x^{00}) dx^0 dx^1 dx^{00}, \tag{5.2}$$

where according to (5.1) the domain of integration is understood as

$$\{t x^0 + \tau(\tfrac{1}{2}t^2 x^{00} - b x^1) > a_n\}. \tag{5.3}$$

Denote by $q(x^0, x^1, x^{00})$ the density function of the limiting normal vector (L^0, L^1, L^{00}), and by $q(x^0), q(x^1, x^{00})$, etc, the corresponding marginal densities. By (3.8) $q(x^0, x^1, x^{00}) = q(x^0)q(x^1, x^{00})$. The second term in the integrand (of order τ^2) contributes $o(\tau^2)$. For within this accuracy the domain of integration can be replaced by $\{t x^0 > a_n\}$, so that the integral of this term factorizes into the product of the integrals w.r.t. x^0 and (x^1, x^{00}). In the second integral x^{00} integrates out, so that this integral becomes

$$\int \left((x^1)^2 - I\right) q(x^1) dx^1 = 0.$$

Hence we can consider only the term $\tau b x^1$ in the integrand of (5.2).

Let $Q_n(y, x^1, x^{00}) = \int_y^\infty q_n(x^0, x^1, x^{00}) dx^0$. Taking into account (5.3) we integrate (5.1) w.r.t. x^0 to obtain

$$\alpha - \alpha' = \iint Q_n(a_n/t - \tau(\tfrac{1}{2}t x^{00} - b x^1/t), x^1, x^{00}) \tau b x^1 dx^1 dx^{00}.$$

Obviously, $\frac{\partial}{\partial y} Q_n(y, x^1, x^{00}) = -q_n(y, x^1, x^{00})$. Thus we have

$$\alpha - \alpha' = A_1 + A_2, \tag{5.4}$$

where

$$A_1 = \tau b \iint Q_n(a_n/t, x^1, x^{00}) x^1 dx^1 dx^{00}, \tag{5.5}$$

$$A_2 = \tau^2 b \iint q_n(a_n/t, x^1, x^{00}) x^1 (\tfrac{1}{2}t x^{00} - b x^1/t) dx^1 dx^{00}. \tag{5.6}$$

Consider A_1. The variable x^{00} integrates out, and returning to the expression of Q_n through q_n we rewrite A_1 as

$$A_1 = \tau b \iint\limits_{\{x_0 > a_n/t\}} x^1 q_n(x^0, x^1) dx^0 dx^1. \tag{5.7}$$

Here we need the one-term Edgeworth expansion for $q_n(x^0, x^1)$. Denote by $\varphi_0(x)$ and $\varphi_1(x)$ the densities of L^0 and L^1, which are the normal densities of $N(0,J)$ and $N(0,I)$. Then

$$\begin{aligned} q_n(y,x) &= \varphi_0(y)\varphi_1(x) \\ &+ \frac{\tau}{6}\left[\ldots + 3E\left((l^0)^2 l^1\right)(-D_y)^2\varphi_0(y)(-D_x)\varphi_1(x) + \ldots\right] + o(\tau), \end{aligned}$$

where D_y and D_x denote the differentiation operators w.r.t. y and x respectively. When substituted into (5.7), the integrals of each term factorize into the products of the respective integrals w.r.t. x^0 and x^1. The main term integrates to zero because one of the factors is the mean value of φ_1. The terms of higher order than τ, obviously, contribute $o(\tau^2)$ into (5.7). The suppressed terms of order τ either vanish or integrate to 0 in (5.7). A generic term in brackets contains the derivatives of φ_0 and φ^1, totally of order 3, multiplied by the corresponding product moment. The term with coefficient $E(l^0)^3$ vanishes since l^0 is odd, while the remaining terms contain the 2nd and 3rd derivatives of $\varphi_1(x)$ giving rise to the 2nd and 3rd order Hermite polynomials orthogonal to x^1 in (5.7) (the first order Hermite polynomial).

We have

$$(-D_x)\varphi_1(x) = I^{-1}(x/\sqrt{I})\varphi(x/\sqrt{I})$$

with $\varphi(\cdot)$ being the standard normal density. Hence we obtain

$$\int x^1(-D_x^1)\varphi_1(x^1)dx^1 = 1.$$

Similarly,

$$\int_{a_n/t}^{\infty}(-D_y)^2\varphi_0(y)dy = (-D_y)\varphi_0(y)\big|_{y=a_n/t} = J^{-3/2}\frac{a_n}{t}\varphi\left(\frac{a_n}{t\sqrt{J}}\right).$$

It is seen from (5.1) that $a_n \to a = t\sqrt{J}u_{1-\alpha}$ and this limiting value can be substituted for a_n with error $o(\tau^2)$. Thus

$$A_1 = \frac{1}{2}\tau^2 bJ^{-1}E\left((l^0)^2 l^1\right)u_{1-\alpha}\varphi(u_{1-\alpha}) + o(\tau^2). \tag{5.8}$$

Now we consider A_2. Since the expression (5.6) contains factor τ^2, we can replace q_n by the limiting normal density q and a_n by a. As we saw,

$$q_n(a/t, x^1, x^{00}) = \varphi_0(a/t)q(x^1, x^{00}).$$

Since $\varphi_0(x) = J^{-1/2}\varphi(x/\sqrt{J})$ we have $\varphi_0(a/t) = J^{-1/2}\varphi(u_{1-\alpha})$.

Hence we obtain from (5.6)

$$A_2 = \tau^2 bJ^{-1/2}\varphi(u_{1-\alpha})\left[\frac{1}{2}tE(l^{00}l^1) - \frac{b}{t}I\right] + o(\tau^2). \tag{5.9}$$

Putting (5.8) and (5.9) into (5.4) we obtain (3.17–18) for a univariate nuisance parameter η. The case of a vector-valued nuisance parameter is treated in a similar way.

5.2 One-term Edgeworth expansion

For the sake of completeness we recall here the formula for the one-term Edgeworth expansion for a statistic admitting a stochastic expansion and present its informal derivation. Suppose we have i.i.d. $p+1$-variate random vectors

$$(Y_{0i}, \mathbf{Y}_i).$$

(Typically we have i.i.d. observations $X_1, \ldots, X_n$ and $p+1$ functions on the range of X's, so that the vector $(Y_{0i}, \mathbf{Y}_i)$ is formed by these functions of X_i.) Denote by $(S_n, \mathbf{T}_n)$ the normalized sums of these vectors,

$$S_n = \tau \sum_{i=1}^{n} Y_{0i}, \quad \mathbf{T}_n = \tau \sum_{i=1}^{n} \mathbf{Y}_i.$$

We are interested in the one-term E-expansion for the distribution of a statistic

$$Z_n = S_n + \tau h(S_n, \mathbf{T}_n),$$

where $h(\cdot)$ is a polynomial of $p+1$ variables.

We assume that Y_{01} has mean zero, variance σ^2 and a finite third moment $\mu_3 = EY_{01}^3$, moreover, its distribution is non-lattice. The moment conditions on the other components and conditions on the joint distribution of the entire vector depend on $h(\cdot)$. The most general moment conditions are given in Chibisov(1980-81); for $h(\cdot)$ a quadratic function (as is the case in (4.20)) a sufficient moment condition on $\mathbf{Y}_1$ is that $Y_{11}, \ldots, Y_{pi}$ have zero means and finite second moments. Assuming this, denote by Σ the covariance matrix of $(Y_{01}, \mathbf{Y}_1)$. Then by the Central Limit Theorem

$$(S_n, \mathbf{T}_n) \xrightarrow{d} (S, \mathbf{T}) \tag{5.10}$$

where $(S, \mathbf{T})$ is a normally distributed random vector in R^{p+1} with mean zero and covariance matrix Σ.

Denote by $F_n(x)$ the d.f. of S_n, $F_n(x) = P(S_n < x)$. Then for $F_n(x)$ the one-term E-expansion holds,

$$F_n(x) = \Phi(x/\sigma) - \tau \frac{\mu_3}{6\sigma^3} H_2(x/\sigma)\varphi(x/\sigma) + o(\tau), \tag{5.11}$$

where $H_2(x) = x^2 - 1$, and $\Phi(\cdot)$ and $\varphi(\cdot)$ denote the standard normal d.f. and density.

To derive the E-expansion for Z_n, assume that S_n has a density $p_n(x)$. Then by the formula for total probability

$$P(Z_n < x) = \int P(\tau h(y, \mathbf{T}_n) < x - y | S_n = y) p_n(y) dy. \tag{5.12}$$

The right-hand side of (5.12) can be rewritten as

$$F_n(x) + \int [P(\tau h(y, \mathbf{T}_n) < x - y | S_n = y) - \mathbf{1}_{(0,\infty)}(x - y)] p_n(y) dy, \tag{5.13}$$

where $\mathbf{1}_A(\cdot)$ denotes the indicator function of the set A. By the change of variables $x - y = \tau z$ the last integral becomes

$$\tau \int [P(h(x-\tau z, \mathbf{T}_n) < z \,|\, S_n = x - \tau z) - \mathbf{1}_{(0,\infty)}(z)] p_n(x - \tau z) dz. \tag{5.14}$$

By (5.10) $p_n(x) \to (1/\sigma)\varphi(x/\sigma)$, and the formal passage to the limit yields that the integral in (5.14) converges to

$$\sigma^{-1}\varphi(1/\sigma) \int [P(h(x, \mathbf{T}) < z \,|\, S = x) - \mathbf{1}_{(0,\infty)}(z)] dz. \tag{5.15}$$

Integrating by parts shows that the integral in (5.15) equals

$$-E(h(x, \mathbf{T}) \,|\, S = x) \tag{5.16}$$

(see Feller (1971), Chapter 5, §6, Lemma 1). Combining (5.11)–(5.16) we obtain

$$P(Z_n < x) = \Phi(x/\sigma) - \tau\varphi(x/\sigma) \Big[\frac{\mu_3}{6\sigma^3} H_2(x/\sigma) +$$

$$+ \sigma^{-1} E(h(x, \mathbf{T}) \,|\, S = x) \Big] + o(\tau). \tag{5.17}$$

5.3 The formula for the power loss

Here we give an informal proof of the formula (2.28). Since parameter t has no special meaning here, we denote the two sequences of probability measures corresponding to hypotheses H_0 and H_1 by $P_{n,0}$ and $P_{n,1}$. We compare the MP test for H_0 against H_1 based on the LLR

$$\Lambda_n = \log \frac{dP_{n,1}}{dP_{n,0}}$$

with the test based on a statistic S_n of the form

$$S_n = \Lambda_n - \tau \Pi_n \tag{5.18}$$

(cf. (2.21) and the notation introduced before (2.27)). The tests reject H_0 for $\Lambda_n > c_n$ and $S_n > b_n$ respectively with

$$P_{n,0}(\Lambda_n > c_n) = P_{n,0}(S_n > b_n) = \alpha \tag{5.19}$$

(see (2.22)). We assume that (Π_n, Λ_n) converges in distribution under $P_{n,0}$ to a nondegenerate bivariate r.v. (Π, Λ) (see (2.27)) and that Λ_n under $P_{n,0}$ has d.f. and density $F_{n,0}$ and $f_{n,0}$ converging to F_0 and f_0. The powers of the two tests are

$$\beta_n^* = P_{n,1}(\Lambda_n > c_n) \quad \text{and} \quad \beta_n = P_{n,1}(S_n > b_n).$$

Using (5.19) their difference can be written as

$$\beta_n^* - \beta_n = E_{n,0}(e^{\Lambda_n} - e^{b_n})(\mathbf{1}_{(c_n,\infty)}(\Lambda_n) - \mathbf{1}_{(b_n,\infty)}(S_n)) = A_n + B_n, \tag{5.20}$$

where

$$A_n = E_{n,0}(e^{\Lambda_n} - e^{b_n})(\mathbf{1}_{(-\infty,b_n)}(\Lambda_n) - \mathbf{1}_{(-\infty,c_n)}(\Lambda_n)), \tag{5.21}$$

$$B_n = E_{n,0}(e^{\Lambda_n} - e^{b_n})(\mathbf{1}_{(-\infty,b_n)}(S_n) - \mathbf{1}_{(-\infty,b_n)}(\Lambda_n)). \tag{5.22}$$

Denote $d_n = c_n - b_n$. We will show that

$$d_n = -\tau E[\Pi|\Lambda = c] + o(\tau), \tag{5.23}$$

$$A_n = -\frac{1}{2} d_n^2 e^c f_0(c) + o(\tau^2), \tag{5.24}$$

$$B_n = \frac{1}{2}\tau^2 e^b E[\Pi^2|\Lambda = c] f_0(c) + o(\tau^2). \tag{5.25}$$

Combined with (5.20) these relations immediately imply (2.28).

Proof of (5.23). Denote the d.f. of S_n by $F_{S_n}(x)$; recall the we denote the d.f. of Λ_n by $F_{n,0}(x)$. Using (5.18) we establish similarly to (5.20)–(5.24)

$$F_{S_n}(x) = F_{n,0}(x) - \tau E[\Pi \mid \Lambda = x] f_0(x) + o(\tau). \tag{5.26}$$

In view of (5.19) this equality implies, in particualar, that $b_n \to c$ and $c_n \to c$ as $n \to \infty$, where c is defined by $F_0(c) = 1 - \alpha$. Put $x = b_n$ in (5.26) and replace $F_{S_n}(b_n)$ by $F_{n,0}(c_n)$ according to (5.19). Then we obtain

$$F_{n,0}(c_n) - F_{n,0}(b_n) = -\tau E[\Pi \mid \Lambda = c] f_0(c) + o(\tau). \tag{5.27}$$

On the other hand,

$$F_{n,0}(c_n) - F_{n,0}(b_n) = d_n(f_0(c) + o(1)). \tag{5.28}$$

Now (5.27) and (5.28) imply (5.23).

Proof of (5.24). Rewrite (5.21) as

$$A_n = e^{b_n} \int_{c_n}^{b_n} (e^{y-b_n} - 1) dF_{n,0}(y).$$

Since $d_n = O(\tau)$ by (5.23), we have

$$e^{y-b_n} = 1 + y - b_n + O(\tau), \;\; y \in [c_n, b_n]$$

and therefore

$$A_n = e^{b_n} \int_{c_n}^{b_n} (y - b_n) f_{n,0} d(y) + o(\tau^2) =$$
$$= -\frac{1}{2} e^c f_0(c) d_n^2 + o(\tau^2),$$

which proves (5.24).

Proof of (5.25). By the formula for total probability rewrite (5.22) as

$$B_n = e^{b_n} \int (e^{y-b_n} - 1)[P(\tau\Pi_n < b_n - y|\Lambda_n = y) - \mathbf{1}_{(-\infty,b_n)}(y)] f_{n,0}(y) dy.$$

By the change of variables $b_n - y = \tau z$ this becomes

$$B_n = \tau e^{b_n} \int (e^{-\tau z} - 1)[P(\Pi_n < z|\Lambda_n = b_n - \tau z) - \mathbf{1}_{(0,\infty)}(z)] f_{n,0}(b_n - \tau z) dz.$$

The conditional distribution of Π_n is essentially concentrated in a bounded domain. Hence $e^{-\tau z} - 1 \approx -\tau z$. By a formal passage to a limit we obtain

$$B_n = -\tau^2 e^c f_0(c) \int z[P(\Pi < z \mid \Lambda = c] - \mathbf{1}_{(0,\infty)}(z)] dz + o(\tau^2).$$

Integrating by parts yields (5.25).

References

Albers, W., Bickel, P. J., and van Zwet, W. R. (1976). Asymptotic expansion for the power of distribution-free tests in the one-sample problem, *Ann. Statist.*, 4, 108–156; correction (1978), 6, 1170–1171.

Bening, V. E. (1995). A formula for deficiency: one sample L-and R-tests. I, II, *Math. Methods of Statist.*, 4, 167–188, 274–293.

Bening, V. E. (1997). Some properties of asymptotically efficient U-statistics, *Teor. Veroyatn. i Primen.*, 42, 382–384 (in Russian).

Bickel, P. J., Chibisov, D. M., and van Zwet, W. R. (1981). On efficiency of first and second order, *International Statistical Review*, 49, 169–175.

Chibisov, D. M. (1980–81). An asymptotic expansion for the distribution of astatistic admitting a stochastic expansion. I, II, *Theory Probab. Appl.*, 25, 732–744, 26, 1–12.

Chibisov, D. M. (1982). Asymptotic expansions in problems of testing hypotheses. I, II, *Izv. Akad. Nauk UzSSR, Ser. Fiz.-Mat. Nauk*, 5, 18–26; 6, 23–30.

Chibisov, D.M. (1983). Asymptotic expansions and deficiencies of tests. In: *Proc. Internat. Congress of Math.*, Warszawa, vol. 2, 1063–1079.

Chibisov, D.M. (1985). Calculation of the deficiency of asymptotically efficient tests, *Theory Probab. Appl.*, 30, 289–310.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications.* vol. 2, Wiley, New York.

Hodges, J. L., Jr. and Lehmann, E. L. (1970). Deficiency, *Ann. Math. Statist.*, 41, 783–801.

Pfanzagl, J. and Wefelmeyer, W. (1978). An asymptotically complete class of tests. *Z. Wahrsch. verw. Gebiete*, 45, 49–72.

Pfanzagl, J. (1980). Asymptotic expansions in parametric statistical theory, In: *Developments in Statist.* (P.R.Krishnaiah ed.), New York, Academic Press, vol. 3, 1–97.